

Data warehouse

A Data warehouse is not a one-time software product or application - it is an important information processing & delivery system architecture for decision making!

Data warehouse Testing - Ensuring Information Quality through Innovation



Data warehousing Testing

When building a Data warehouse, organizations employ either a top-down or bottom-up development approach. In the top-down approach, an enterprise data warehouse (EDW) is built and information processes are created business area by business area with a underlying dependent data marts pulled out from the EDW contents. In the bottom-up approach, independent data marts with information processes are created with the view to integrating them into an EDW in the future. Either way, a major blind spot in the data warehousing project delivery quality is testing. Data warehouse testing may have the same principles/fundamentals of a general testing project, but testing DWH & BI projects involves significant programming work as there are limited front end screens but mostly back end processes that work on data sets.

Organisations do not have enough time to do a complete testing job as their operations job takes precedence over testing work. They also lack testing skills and perspective to carry out testing, more so in Data warehouse testing. The attitude and methodology required for DWH testing are not same as those required for normal testing. The skill sets required include resources with a strong aptitude and with technology specialization and programming knowledge.

Challenges

Some of the data challenges that impact the business of an organisation are listed below.

- Loss of data during the ETL process
- Inaccurate and incomplete data coming from disparate source systems
- Data duplication appearing in source feeds
- Differing definitions of operational data from disparate systems.
- Authentication and security concerns
- Multiple country integration issues, business definition and source systems disparity
- Non-availability of comprehensive test bed
- Imbalances in performance across the different environment like Development, UAT and Production
- Data Quality Issues

DWH Testing attempts to plug each of the above gaps by ensuring that every stage of the source to target movement/ transformation of data is tested and is working fine.

DWH Testing Explained ...

- **Verification of data elements**
- **Database procedures are used to validate**
 - Mappings
 - Transformations
 - Dimensions
 - Checksum
- **Prepare Integrity SQL's to verify**
 - Hash Count & Total
 - Standard Aggregates
 - Standard Analytics
 - Dimension Data
 - Transformation Data
 - Low Cardinality
 - Derived Attributes
 - Data Reconciliation
 - Integrity
- **Random checks on data based on sampling**
- **Access control**

Gartner research indicates that the cost of poor data quality to North American corporations as shown by documented studies to be around \$600bn!



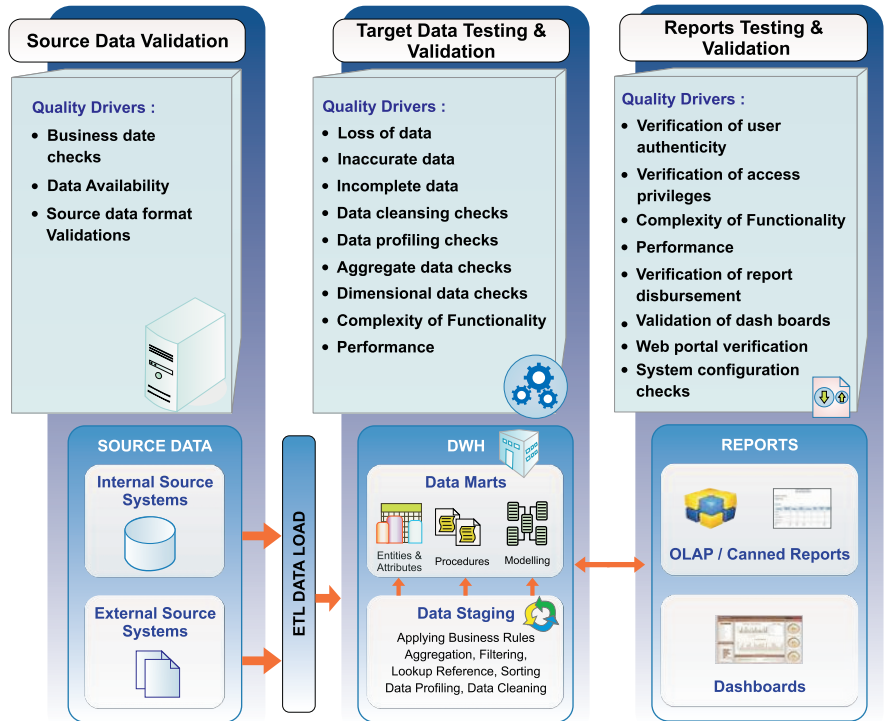
■ Data warehouse testing approach

Acuma's DWH testing approach called "3 segments & 3 steps", is broadly divided into ETL, DWH and Report Segments driven by these quality drivers, shown in the picture on the right. In each segment we approach testing in 3 steps:

Step 1: This includes simple data extraction and validation as well as business rules and functionalities

Step 2: This includes testing of primary / foreign / surrogate key references, data mapping and data transformations

Step 3: This includes error handling and tracking and performance at an individual table or data level



■ Regression Testing

In today's business scenario, the demand and use of BI applications and applications based on warehouses has grown by leaps and bounds and shows no signs of slowing down. Applications are undergoing changes faster than they can be tested. In this context, Regression testing is gaining more and more importance as data quality is now of paramount importance.

Regression testing is revalidation of existing functionality with every new release of the code. The goal of regression testing is to ensure that a change to the system hasn't broken existing functionality. Therefore effective regression testing is a key enabler for successfully enhancing value of any Information Management solution.

■ Organization

Acuma is Saksoft's fully owned subsidiary. Our solutions are a combination of domain knowledge and technical expertise. We build solutions that follow the best industry practices and deliver IT solutions that enhance your business value. Our services and products are modeled to aid corporates in identifying their pain points and providing solutions which will use available information to improve processes, maximize profitability and manage risk better.

Saksoft meets the requirements of CMMI level 5 accredited quality system which helps it to deliver predictable quality solutions.

■ Assure

Assure is a complete V&V service offering from Acuma. It not only verifies that your application/product is working, but also validates that it is the right application/product for the expected business use, thus ensuring that it is deployable and useful. This also helps an organisation to easily deploy new processes or change existing ones.

We provide complete cycle of testing services

- Test Strategy Definition
- Test Plan Creation
- Test Case Writing (Manual and Automated)
- Test Execution (Manual and Automated) and
- Certification (that requirements have been met).